DOCUMENT RESUME

ED 046 895 24 SP 004 643

AUTHOR	Pinney, Robert H.
TITLE	Teacher Presentational Behaviors Related to Student Achievement in English and Social Studies.
INSTITUTION	Stanford Univ., Calif. Stanford Center for Research and Development in Teaching.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO	TR-16
BUREAU NO	BR-5-0252
PUB DATE	Dec 70
CONTRACT	OEC-6-10-078
NOTE	47p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Analysis of Variance, *Behavioral Science Research, *Behavior Patterns, *Behavior Theories, Educational Research, Effective Teaching, *Statistical Analysis

ABSTRACT

This study attempted to identify correlates of successful teaching in the presentation behavior of pre-intern teachers. The sample was drawn from 54 Stanford students who each taught one of six preset social studies or English lessons to 25 eighth and ninth graders. Each pupil took a multiple-choice comprehension test which was used to compute a mean class achievement score for each teacher. The score was adjusted twice, by analysis of covariance for student verbal and quantitative ability, and also for lesson difficulty. These scores were used to select 16 high-scoring and 16 low-scoring teachers. Four categories of presentation behavior (verbal, nonverbal, combination, and interaction) were investigated, yielding more than 100 measures of behavior. Audiotape, videotape, or typewritten transcript records were used to observe and code behaviors. Statistical analysis of each subsample and the total sample were conducted using analysis of variance, rank correlation, stepwise discriminate analysis, stepwise regression analysis, and sign test procedures. The results indicated that the behaviors related to success in one sample differed from those in the other; few behaviors discriminated high-scoring from low-scoring teachers when the two samples were combined. It seemed that high-scoring teachers conveyed the essential points by frequent use of repetition, verbal statements of importance, and/or reinforcement of pupil responses. (Author/MBM)

ED046895

STANFORD CENTER
FOR RESEARCH AND DEVELOPMENT
IN TEACHING

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Technical Report No. 16

TEACHER PRESENTATIONAL BEHAVIORS RELATED TO STUDENT ACHIEVEMENT IN ENGLISH AND SOCIAL STUDIES

Robert H. Pinney

School of Education
Stanford University
Stanford, California

December 1970

Published by the Stanford Center for Research and Development in Teaching, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the Office of Education. (Contract No. OE-6-10-078, Project No. 5-0252-0501.)

SP004643

Introductory Statement

The central mission of the Stanford Center for Research and Development in Teaching is to contribute to the improvement of teaching in American schools. Given the urgency of the times, technological developments, and advances in knowledge from the behavioral sciences about teaching and learning, the Center works on the assumption that a fundamental reformulation of the future role of the teacher will take place. The Center's mission is to specify as clearly, and on as empirical a basis as possible, the direction of that reformulation, to help shape it, to fashion and validate programs for training and retraining teachers in accordance with it, and to develop and test materials and procedures for use in these new training programs.

The Center is at work in three interrelated problem areas: (a) Heuristic Teaching, which aims at promoting self-motivated and sustained inquiry in students, emphasizes affective as well as cognitive processes, and places a high premium upon the uniqueness of each pupil, teacher, and learning situation; (b) The Environment for Teaching, which aims at making schools more flexible so that pupils, teachers, and learning materials can be brought together in ways that take account of their many differences; and (c) Teaching Students from Low-Income Areas, which aims to determine whether more heuristically oriented teachers and more open kinds of schools can and should be developed to improve the education of those currently labeled as the poor and the disadvantaged.

Technical Report No. 16, which follows, was completed under the program on Heuristic Teaching as part of the Training Studies project. Data from the Stanford Intern Data Bank were used to study four broad categories of teacher presentation behavior: verbal, nonverbal, combined verbal/nonverbal, and teacher-student interaction. Compared with teachers of low-scoring students, teachers of high-scoring students did a better job of conveying the essential points of the lesson by emphasizing them through repetition, verbal statements of importance, and reinforcement of pupil responses.

Table of Contents

	Page
Abstract	vii
Introduction	1
Relating Teacher Behavior to Student Behavior	1
Explaining as a Teaching Task	3
Rationale	5
Procedures	6
Suitability of the Intern Data Bank	8
Adjusting Student Achievement Scores for Sample Selection	10
Behaviors Investigated	11
Statistical Methods	12
Results	12
Verbal Behaviors	12
Nonverbal Behaviors	17
Combinations of Behaviors	22
Results of Stepwise Discriminant and Stepwise Regression Analysis	25
Summary	33
The June Subsample	33
The August Subsample	35
The Total Sample	36
Paucity of Significant Variables	37
Implications for Teaching	39
Implications for Future Research	39

Abstract

This study attempted to identify correlates of successful teaching in the presentation behavior of pre-intern teachers. Presentation was defined as the verbal and nonverbal behaviors used by the teacher when presenting previously organized content.

The sample was drawn from 54 Stanford University pre-interns who each taught one of six preset social studies or English lessons to 25 eighth- and ninth-grade students in June 1967. Each student took an appropriate 20-item multiple-choice comprehension test which was used to compute a mean class achievement score for each teacher. This score was adjusted for student verbal and quantitative ability, using analysis of covariance, which yielded a once-adjusted mean class achievement score for each teacher. To adjust for lesson difficulty, the median once-adjusted score for each lesson was subtracted from each teacher's once-adjusted score, yielding a twice-adjusted mean class achievement score for each teacher.

These twice-adjusted scores were used to select the eight high-scoring and eight low-scoring teachers to form a June subsample. A similarly selected group was drawn from the August 1967 population (August subsample). These two subsamples were combined to form a total sample of 16 high-scoring and 16 low-scoring teachers.

Four categories of presentation behavior (verbal, nonverbal, combination, and interaction) were investigated, yielding more than 100 measures of behavior. Coding procedures were established for each measure, and audiotape, videotape, or typewritten transcript records of teacher performance were used to observe and code behaviors.

Statistical analyses of each subsample and the total sample were conducted using analysis of variance, rank correlation, stepwise discriminant analysis, stepwise regression analysis, and sign test procedures.

Of the behaviors measured, three achieved F-ratios significant at the .05 level or greater. These were verbal markers of importance,

verbal markers of importance used in proximity to either distributed or massed repetition, and percent of multiply reinforced student responses.

Although interpretations must remain speculative, the results indicate that the behaviors related to success in one sample differed from those in the other; few behaviors discriminated high-scoring from low-scoring teachers when the two samples were combined. Despite these differences, it seems that the high-scoring teachers conveyed the essential points of the lessons by emphasizing them through frequent use of repetition, verbal statements of importance, and/or reinforcement of pupil responses. These verbal emphases were often accompanied by various nonverbal behaviors.

TEACHER PRESENTATIONAL BEHAVIORS RELATED TO STUDENT ACHIEVEMENT IN ENGLISH AND SOCIAL STUDIES

Robert H. Pinney[1]
Upper Midwest Regional Educational Laboratory

This study represents an attempt to describe behaviorally the ways in which successful teachers differ from less successful teachers when confronted with the task of "explaining," specifically those teacher behaviors subsumed by the term "presentational variables." Presentational variables deal with those verbal and nonverbal behaviors manifested by the teacher in the process of presenting the content of a given lesson, as opposed to those dealing with the organizing, sequencing, and structuring of the content of a given lesson.

Relating Teacher Behavior to Student Behavior

Summaries of the literature conclude that only limited attempts have been made to relate specific teacher behaviors or patterns of behavior to specific criteria for student achievement (e.g., Medley & Mitzel, 1959, 1963; Bellack & Davitz, 1963; Taba, 1964; Gallagher & Aschner, 1963; Smith & Meux, 1959; Smith, 1963). In the majority of the studies reviewed, the primary criterion (dependent variable) has been a global rating of "teacher competence," usually assigned to the teacher by a superordinate. The correlates of such measures of effectiveness have usually been personality traits, characteristics, and generic instructional strategies displayed by the teacher (Ryans, 1960).

Although the correlations resulting from these studies have been numerous and interesting, the overall results have been disappointing (Gage, 1963, p. 118). The results have too often been nonsignificant, inconsistent, and lacking in substantial psychological and educational meaning.

---

[1]Dr. Pinney was a Research Assistant at SCRDT when this study was carried out.

A rationale often cited for this lack of concrete results is that teaching is a complex task, and that any psychometric approach to the measurement of teaching which assumes common and stable factors in the teacher and situation must result in indifferent success. Teaching success is multidimensional, and there are many kinds of teachers, programs, and situations (Ackerman, 1954; Mitzel & Gross, 1958).

Recent attempts to increase the amount of dependable and meaningful information about teacher behavior correlates have in common the characteristic of decreased complexity. That is, some researchers are adopting at least the spirit of what Gage (1963) refers to as "microcriteria of effectiveness":

> Rather than seek criteria for the overall effectiveness of teachers in the many, varied facets of their roles, we may have better success with criteria of effectiveness in small, specifically defined aspects of the role. Many scientific problems have eventually been solved by being analyzed into smaller problems whose variables were less complex (p. 120).

Approximating the notion of microcriteria are the recent and quite different efforts of Flanders (1960) and Bellack, Kleibard, Hyman, & Smith (1966). Although each approached the observation and coding of teacher-pupil classroom behavior differently, each attempted at the outset to narrow the complexity of the study by (a) reducing the criteria of effectiveness to student achievement over a one- or two-week period, (b) using more specific criterion tests, and (c) providing teachers with standardized and previously unused teaching materials.

Using a high degree of specificity, Gallagher and Aschner (1963) reported that a high frequency of divergent questioning by the teacher was correlated with a high frequency of divergent thinking productivity among children. According to Berlyne (1966), prequestioning is related to curiosity and retention. In the area of lesson presentation, Coats and Smidchens (1966) found speaker dynamism to be significantly related to success on tests of immediate recall.

Thus, the past five to ten years have been marked by increased attempts to assess the consequences of specific dimensions of teacher behavior on specific dimensions of learner behavior.

Explaining as a Teaching Task

Although the term "explaining" or "explanation" is used frequently in educational literature and discussions, its meaning is inexact and changes from situation to situation. It seemed unnecessary to become involved in philosophical discussions of the term, attempting to differentiate between "genetic explanation," "teleological explanation," "historical explanation," and "causal explanation" (Swift, 1961). Rather, the definition of explaining put forth by Thyne (1963) seems most appropriate. That is, explanation may be considered as "any procedure which results in understanding" (p. 126). This implies that not all explaining requires talking and not all talking is explaining. This is to say that unless understanding takes place, explaining has not occurred. The successful explainer is the one who selects the most important cues and so restructures them as to produce understanding of the relationships between a problem and its solution, a cause and its effect, or an antecedent condition and consequent results.

There is considerable research literature dealing with the problem of isolating teacher behaviors related to effectiveness in explaining. Much of this research has been conducted by investigators trained in the fields of public speaking (Thompson, 1967; Petrie, 1963; Beighley, 1954). A psychologist, Rudin (1961), revealed significant variance in teacher effectiveness in a science-lecture situation. His results provide no information as to the teachers' behavioral differences or to what extent these differences in effectiveness might be generalizable over topics, students, and time.

In an attempt to throw light on the question of generality of effectiveness in explaining, Fortune, Gage, and Shutes (1966) conducted a study in which effectiveness in explaining was determined by student performance on a short criterion test. Limiting the task, time, and content, they attempted to test the generality of explaining effectiveness over topics, students, and both.

The results indicated that effectiveness in explaining was fairly general only over groups of students (median r = .40), and not over topics (median r = .00).

In a similar attempt, Belgard, Rosenshine, and Gage (1967) collected data on 43 high school social studies teachers. Each teacher presented two lectures to his twelfth-grade class, one on Yugoslavia and one on Thailand. Each lecture was videotaped. All teachers were restricted to a lecture-only format and to the material provided by the experimenters. A 10-item multiple-choice test was administered following each lecture. The mean class score for each teacher was obtained for both lessons and adjusted both for student ability, as measured by the score on a third lesson, and for content relevance of the lecture.

After adjusting mean class scores for student ability and content relevance, Rosenshine (1968) used the resulting residual mean scores to rank the teachers from "high" to "low" in explaining effectiveness. The 100 or more variables initially investigated by Rosenshine were derived from four areas of research, namely linguistics, instructional set, speech presentation, and teacher behavior.

Four variables were found to differentiate at the .05 level between high and low teachers. The first two of the four discriminating variables are essentially organizational, and the other two are essentially presentational, the latter being the focus of the present study. The first discriminating variable, "explaining words and phrases," consists of the frequency of words and phrases

which answer questions such as "How?" "Why?" and "With what consequences?" The second variable, a "rule-example-rule" pattern of organization, consists of the frequency with which the teacher sequences rules and examples, in a "rule-example-rule" sequence as compared to "rule-example" and "example-rule" sequences. A third variable, "gestures," is a presentational variable and was found to be a significant discriminator when reported in terms of "total number of gestures per lecture." The fourth variable, "movement," is a presentational variable that deals with the frequency with which the teacher moves from right to left about the room.

Although these results are subject to several interpretations, such is not the purpose of this report. Suffice it to say that certain features of the data, including its collection and analysis, are related to the reported outcomes.

Rationale

The kind of investigation reported above is useful. It allows us to investigate possible correlates of effective teaching, yet supports the bias of most educators that there are probably few such discrete and independent teaching behaviors that can be used to predict teacher effectiveness. The purpose of the present study is to utilize the above findings in an effort to refine and extend the generalizability of the results.

First, it seems important to test the findings of previous studies under similar, yet somewhat different, circumstances. As a form of replication, this may promote a more definitive interpretation of previous findings.

Second, removing certain limitations, while adding certain controls, will promote the analysis of additional behaviors and combinations of behaviors for their relation to effective teaching.

Third, to achieve greater specificity and flexibility of analysis, this study is directed toward only one broad dimension

of teacher behavior, presentational behavior. Such a strategy allows one to approach each promising variable from a variety of positions rather than from a "one-only" position. The freedom to redefine, combine, and arrange various sequences of teacher behavior might lead to the delineation of more significant behaviors related to effective teaching.

Since significant is a somewhat nebulous term, its relevance to this investigation should be discussed. As used here, significant refers to the degree to which a particular variable can be made significant, rather than to its inherent semantic significance. It is possible then to think of the significance of a particular behavior in terms of the function it serves. The teacher, as a primary mediator of learning in the classroom, is seen as being faced with a specific task: to explain (promote understanding). The task of explaining has at least two subtasks: (a) to structure (impose some organization upon) the content, and (b) to present (transact) this structured content. In so doing, the teacher will behave in ways necessary to the fulfillment of functions germane to each of the subtasks. If it can be assumed that functions are describable and definable with varying degrees of generality, one research task becomes that of posing questions regarding the probable behaviors or classes of behaviors that serve certain functions.

Procedures

This study utilized data collected by the Stanford Center for Research and Development in Teaching as part of the Intern Data Bank. The goal of the Intern Data Bank was to establish an extensive pool of data on all interns in the Stanford Secondary Teacher Education Program (STEP). These data were collected at three times: June 1967, August 1967, and May 1968. Although the project resulted in the collection of a wealth of test, appraisal, inventory, and biographical information, only that for the 54 interns who taught preset lessons in June and August 1967 was used in this study. These interns were enrolled in the microteaching clinic.

Prior to June 1967, six preset lessons were developed by staff members working with the Center and STEP, including Prof. F. J. McDonald, James Cooper, and Earl Seidman. Each of the lessons, four in social studies and two in English, was composed of a list of the lesson's major objectives, a set of illustrative materials, and a 20-item multiple-choice comprehension test. Fron the total population of social studies (N = 65) and English (N = 41) interns, 36 social studies and 18 English interns were randomly selected. Each of the randomly selected social studies interns was randomly assigned to teach one of the four social studies preset lessons, and each of the randomly selected English interns was randomly assigned to teach one of the two preset English lessons. Each of these 54 interns was then randomly assigned to a specific class of students.

The student population was representative of a suburban area, which included families ranging from high to low socio-economic status. All students had just completed either the eighth or ninth grade. Each student was assigned a code number and randomly assigned to one of 20 class groups, the groups averaging 22 students in size.

Thus, 54 randomly selected social studies and English interns were randomly assigned to teach one of six preset social studies or English lessons to a class of 20-25 randomly assigned post-eighth- and post-ninth-grade students at a randomly assigned time and place. Except for the administration of ability tests to the students, the entire June administration was completed in three consecutive days.

At the end of each preset lesson, students were given the 20-item multiple-choice comprehension test which required about 20 minutes. These tests were administered by graduate student proctors.

All interns repeated this sequence in August, at the end of the summer microteaching clinic. Those interns who taught preset lessons in June also taught preset lessons in August, but their lessons and students were different. Student ability tests were also administered.

Suitability of the Intern Data Bank

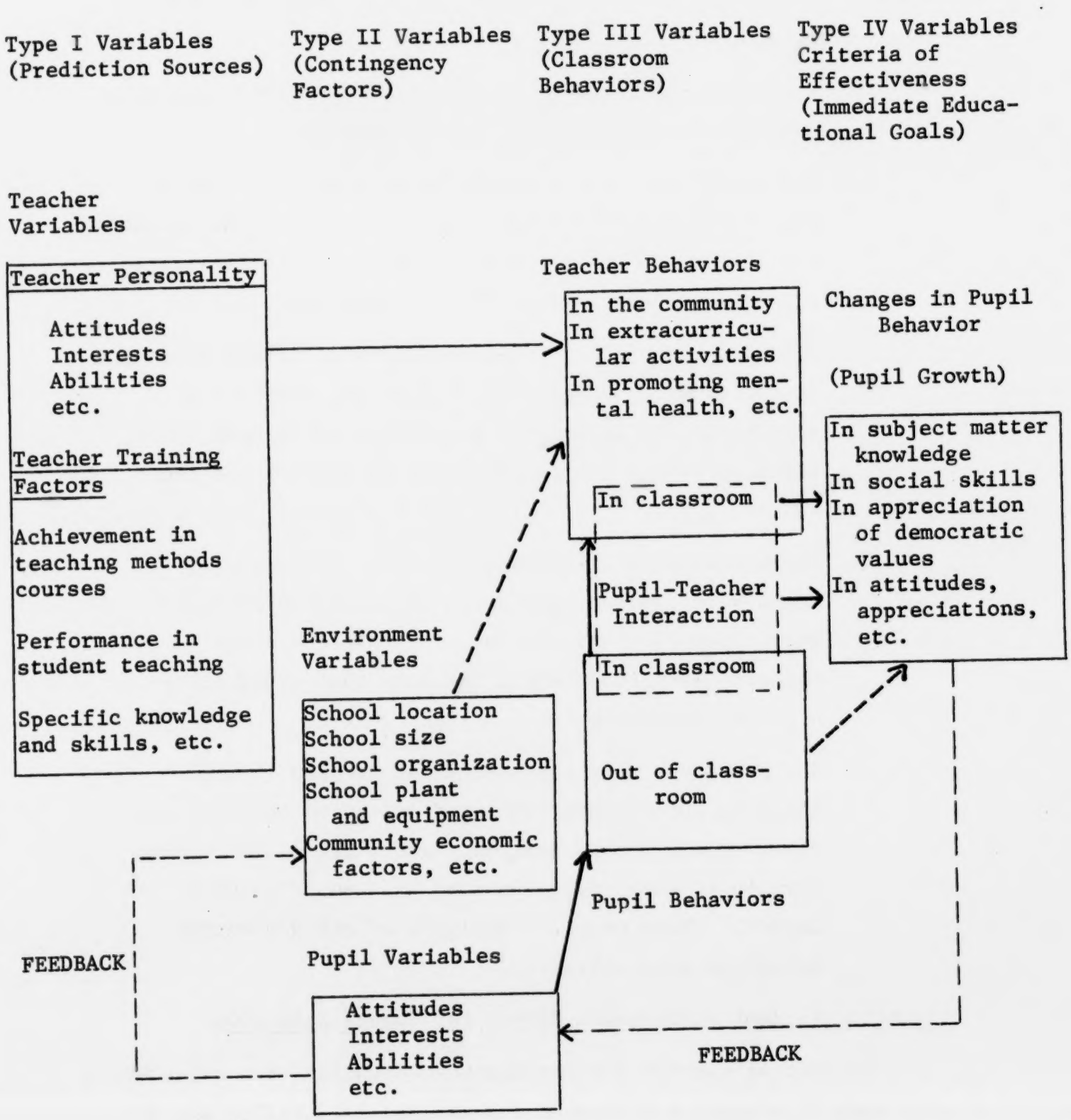
These procedures yielded data suitable for the types of analyses proposed here. Obviously, the almost complete randomization procedures control for almost every threat to the internal validity of the results obtained, although external validity (generalizability) is limited by the nonrandom selection of interns and students. As stated by Campbell and Stanley (1963, p. 195), "the most adequate all-purpose assurance of lack of initial biases between groups is randomization. Within the limits of confidence stated by the tests of significance, randomization can suffice without the pretest." The Generalized Schema for Research in Teacher Effectiveness (Mitzel, 1957, p. 5) provides a means by which to discuss the various controls used in the Intern Data Bank (Figure 1).

Among Type I Prediction Variables the two major factors outlined by Mitzel are "teacher personality" and "teacher training." In the present study, teacher personality factors were randomly distributed throughout the sample. Teacher training factors were well controlled since the data were collected before any teacher training or professional course work had begun. In a sense, all subjects were equal in their "knowledge of good teaching."

Beyond these controls over teacher variables, the greatest value of the Intern Data Bank comes from the extensive control maintained over what Mitzel refers to as Type II Variables (Contingency Factors). These factors consist of environmental and pupil variables. The general situation under which the data were collected held constant across interns such factors as school location, school size, school organization, school plant and

FIGURE 1

Generalized Schema for Research in
Teacher Effectiveness (Mitzel, 1957, p. 5)



equipment, community economic factors, etc., in the project. All data were collected at one school, in a period of three days, with one student population, in classrooms with the same size, shape, furniture, and videotape recording equipment.

Other characteristics of these data that made them suitable for the proposed investigation were the following:

1. The number of lessons taught (N = 54 in both June and August for a total of 108) was adequate for the selection of the extremely high-scoring and extremely low-scoring lessons, as described in the next section.

2. The fact that 12 different lessons were taught (eight in social studies and four in English) provided an opportunity to assess the generality of factors in teaching effectiveness over several lessons and two subject areas.

3. Because students, lessons, teachers, and teaching times were randomly assigned, many rival hypotheses were controlled and the investigator could make certain assumptions about the data that would otherwise be impossible.

4. The permanent videotape recording of each intern's teaching performance permitted extensive analysis of certain behaviors. Written transcripts of the audio portion of each videotape were available for each lesson. These made the analysis of certain verbal behaviors more reliable.

Adjusting Student Achievement Scores for Sample Selection

To control further for contingency variables, two adjustments were made of student achievement scores, one for ability and one for lesson difficulty.

Each class's mean achievement score was then adjusted for student verbal and quantitative ability, yielding a "once-adjusted score" for each class (teacher). To adjust for the difficulty of each lesson, the median of the once-adjusted mean scores was computed for each lesson and subtracted from each once-adjusted score, yielding a twice-adjusted score stated in terms of deviation from the lesson median.

The teachers were ranked according to their twice-adjusted class mean achievement score, and the eight teachers with the highest ranks and the eight teachers with the lowest ranks were selected to form the June subsample. On the basis of the twice-adjusted scores from the August lesson, a similar subsample of 16 teachers was selected. These two subsamples produced a total sample of 16 high-scoring and 16 low-scoring teachers, of whom 24 taught social studies lessons and eight taught English lessons.

Behaviors Investigated

As used in this study, teacher presentation behaviors are defined as those nonsubstantive and nonorganizational characteristics of the teacher's behavior which might be relevant to or affect student understanding of lesson content. Such behaviors (verbal and nonverbal) were investigated to determine whether they were related to student understanding.

Four broad categories of presentation behavior were investigated, each consisting of several classes of behavior, and each class consisting of several individual measures. The four categories were verbal, nonverbal, combinations, and interaction.

Procedures were developed for coding each category, class, and individual behavior. Depending on the behavior being coded, the coders used videotape, audiotape, or typewritten transcript records of the teacher's behavior. The procedure included the definitions and examples for each behavior, as well as instructions for coders. Measures of each behavior were computed in terms

of frequency, frequency per minute, frequency per minute of teacher talk, or percent of occurrence with other behaviors.

Statistical Methods

The statistical methods consisted of rank-order correlation with level of effectiveness, analysis of variance between the 16 high and 16 low teachers, stepwise discriminant analysis, stepwise regression analysis, and the sign test. All of these methods except the sign test were applied to both subsamples as well as the total sample. These statistical procedures allowed the investigator to observe the consistency of various findings across the two subsamples as well as aid in determining the effect of each subsample upon the results for the total (combined) sample.

Results

Of the various measures of teacher behavior gathered in this study, several achieved levels of significance that make them worthy of review and extended discussion. Since these results were obtained through a correlational (rather than experimental) analysis of numerous behaviors, the following conclusions and interpretations must necessarily be tentative. The dual needs for replication and experimental testing are obvious.

Verbal Behaviors

As shown in Table 1, differences between high and low teachers in three measures of teacher verbal behavior proved consistently significant at the .05 level or better. These were:

1. Verbal markers of importance (examples: "The important point to remember...," "Be sure to remember that...."
2. Verbal markers of importance used in proximity to either distributed or massed repetition (examples: "In summary, the three important

points are...," "As I said earlier, the important difference between A and B is...."

3. Percent of multiply reinforced student responses.

The first two are related behaviors and achieved the most consistent and significant results for both subsamples of any behavior investigated. The combined results of these two measures seemed to indicate that teachers--when faced with a preassigned lesson, test, and time period, and with new students--achieve their objectives more effectively when certain key points, principles, facts, etc., are consistently highlighted during the course of the lesson. As indicated by their titles, these two teacher behaviors are verbal statements which highlight, by announcement or repetition, the importance of previous or forthcoming words, phrases, or statements.

The third significant verbal behavior measure, multiply reinforced student responses, seems to function in much the same way, and may indeed be a measure of the same behavior exhibited under different circumstances. This relationship to verbal emphases is clearer when we note that the most frequent form of multiple reinforcement is one of following a student response with a simple reinforcement, and then either a repetition of the response or a statement as to its importance. For example:

P: The Mint Ratio affects the Market Ratio.

T: Very good. The Mint Ratio affects the Market Ratio.

or

That's a very important point.

or

That's the point to remember.

TABLE 1

Mean, Standard Deviation, Correlation With Twice-Adjusted Achievement Scores, Analysis of Variance, and Sign Test Statistics for all Significant Verbal Behaviors

Behavior by Sample	Mean H	Mean L	SD H	SD L	rho	F	Sign[a]
1. Verbal Markers of Importance							
June Subsample (N=16)	.37	.22	.11	.10	.56	7.9*	
August Subsample (N=16)	.39	.19	.25	.08	.47	4.3*	
Total Sample (N=32)	.38	.21	.19	.08	.47	10.6**	.01
2. Verbal Markers of Importance Used in Proximity to Distribution or Massed Repetition							
June Subsample (N=16)	3.00	1.38	1.60	1.51	.45	4.4*	
August Subsample (N=16)	3.00	.50	3.55	.76	.47	3.8	
Total Sample (N=32)	3.00	.94	2.66	1.24	.46	7.9**	.17
3. Percent Distribution or Massed Repetition Used with Verbal Markers of Importance							
June Subsample (N=16)	.27	.13	.17	.10	.42	4.0	
August Subsample (N=16)	.26	.06	.26	.08	.53	5.3*	
Total Sample (N=32)	.27	.10	.21	.09	.48	9.5**	.17

TABLE 1 (Continued)

Behavior by Sample	Mean H	Mean L	SD H	SD L	rho	F	Sign[a]
4. Massed Repetition							
June Subsample (N=16)	.05	.01	.04	.02	.56	6.4*	
August Subsample (N=16)	.04	.03	.04	.04	.13	0.1	
Total Sample (N=32)	.04	.02	.04	.04	.32	3.1	NS
5. Percent of Multiply Reinforced Student Responses							
June Subsample (N=16)	.16	.13	.12	.12	.13	0.3	
August Subsample (N=16)	.27	.15	.10	.08	.56	7.0*	
Total Sample (N=32)	.22	.14	.12	.10	.36	4.2*	.17

H = High-scoring

L = Low-scoring

* $p < .05$

** $p < .01$

[a] Within lessons (N = 10)

Thus, most multiply reinforced student responses tend to emphasize words, phrases, and statements in much the same manner as verbal markers of importance or repetitions.

Although the various measures of probing and reinforcement behavior did not achieve statistical significance, the results are worthy of some discussion. First, it should be noted that each teacher was given approximately four days of intensive training in the use of each of these two behaviors during the summer microteaching clinic. Consequently, the frequency of these two behaviors in both the high- and low-scoring groups was greater in the August sample than in the June sample. The frequency with which these behaviors occurred had no significant relation to effectiveness, and indeed many of the correlation coefficients became negative in the August sample. The only significant measure was percent of nonreinforced student responses, which, when compared to the "almost significant" measure of verbal punishment, would seem to indicate that more frequent statements of verbal punishment, and failure to reinforce positively a greater proportion of pupil responses are negatively related to effectiveness.

Of further interest is the fact that the use of probing and reinforcement was hypothesized to be related to increases in student participation. The mean frequency of student participation is approximately 50 percent greater for both high- and low-scoring groups in the August sample than in the June sample. Yet, the correlation with effectiveness is negative in the August sample, and near-zero in the June sample. Also, the intercorrelations between frequency of student participation and the various measures of probing and reinforcing behavior were very high and positive in the June sample, but very low and negative in the August sample. Although inferences must remain tentative, it can be hypothesized that there is a linear relationship between the frequency of probing and reinforcing behaviors and the frequency of student participation,

but that there is a point at which increases in the latter are not related to increases in the former.

In short, the results obtained in the verbal category of behaviors are neither surprising nor alarming. The findings merely say that teachers who offer explicit verbal cues to the salient points of a lesson will gain greater student recall of those points. Any variation in the amount of questioning or reinforcing behavior seems irrelevant to success in a brief teaching encounter. Although the above discussion is speculative, the basic results are interesting enough to warrant their being investigated in greater detail and with greater control over other variables.

Nonverbal Behaviors

In the nonverbal category, four classes of visual and vocal behavior were investigated: gestures, movements, facial expressions, and vocal proficiency. These classes yielded 47 measures.

Gestures. Significant variables are summarized in Table 2. Only two measures of teacher gesturing behavior achieved statistical significance; total nonfocused congruent gestures and total noncongruent gestures. These results are somewhat contrary to those expected and hence are slightly perplexing. However, some understanding can be gained by considering other factors that may have affected the results.

First, in the matter of congruency, it should be remembered that congruent gesturing is defined as the simultaneous use of a gesture with one or more selected verbal behaviors. Thus, the frequency of a congruent behavior is a function of the frequency with which the two separate behaviors occurred. Teachers who used any of the verbal behaviors (or the gesture) more frequently than others would most likely have a higher frequency of the use of the two when combined. Since many of the selected verbal behaviors had high negative correlations with success in the June, August, and combined samples, one would expect any combination of one or more of those behaviors also to have a negative correlation. Furthering

TABLE 2

Mean, Standard Deviation, Correlation With Twice-Adjusted Achievement Scores, Analysis of Variance, and Sign Test Statistics for all Significant Nonverbal Behaviors

Behavior by Sample	Mean H	Mean L	SD H	SD L	rho	F	Sign[a]
1. Vocal Intensity							
June Subsample (N=16)	4.50	3.25	.93	1.03	.54	6.5*	
August Subsample (N=16)	4.50	3.88	.76	1.13	.35	1.7	
Total Sample (N=32)	4.50	3.56	.82	1.09	.44	7.6**	.02
2. Total Vocal Score							
June Subsample (N=16)	13.00	10.38	2.14	3.02	.43	4.0	
August Subsample (N=16)	12.50	11.63	2.20	2.88	.21	0.5	
Total Sample (N=32)	12.75	11.00	2.11	2.92	.32	3.8	.03
3. Noncongruent Gestures							
June Subsample (N=16)	4.89	3.48	1.25	1.11	.51	5.7*	
August Subsample (N=16)	4.04	3.28	1.79	1.75	.18	0.7	
Total Sample (N=32)	4.46	3.38	1.55	1.42	.25	4.3*	.05
4. Nonfocused, Congruent Gestures							
June Subsample (N=16)	1.74	2.23	.60	.66	-.37	2.4	
August Subsample (N=16)	1.70	2.36	1.04	1.05	-.32	1.6	
Total Sample (N=32)	1.72	2.29	.82	.85	-.34	3.4	.01

TABLE 2 (Continued)

Behavior by Sample	Mean H	Mean L	SD H	SD L	rho	F	Sign[a]
5. Congruent Facial Expressions							
June Subsample (N=16)	.45	.64	.51	.71	-.16	0.4	
August Subsample (N=16)	.33	.83	.37	.50	-.51	5.1*	
Total Sample (N=32)	.39	.73	.44	.60	-.34	3.4	.01

H = High-scoring

L = Low-scoring

* $p < .05$

** $p < .01$

[a] Within lessons (N = 10)

this probability is the fact that the verbal behaviors that had the negative correlations with success were also those that appeared with the highest frequency in both groups. For example, probes, reinforcements, and inquiry questions each occurred about three times as frequently as verbal markers of importance, repetition, or teacher-answered questions.

Since the frequency of nonfocused gestures has an almost consistent zero correlation in all three samples, it stands to reason that the frequency of the combined behavior would be related to the frequency of the verbal behaviors that determine congruency. Thus, the real significance of the negative correlation between nonfocused congruent gestures and effectiveness, and the positive correlation between noncongruent gestures and effectiveness is reduced by the factors controlling their frequency of occurrence.

In the case of focused vs. nonfocused gestures, the results slightly favor the use of focused gestures. However, the differences are so slight as to make unwarranted any speculation about the relative efficacy of the two forms of gesturing. The disturbing fact is not that this distinction shows no clear-cut result, but rather that total gestures showed no significant correlation with effectiveness in this study, although the correlations were all positive.

So many studies of speech delivery and teacher behavior have indicated that the use of gestures and other forms of visual animation are related to greater levels of student achievement, that one would have expected it to be true with these 32 teachers as well. Yet, the results here indicate not only that the number of focused gestures may make no difference, but also that the total number of gestures may make no difference. In short, one must almost revert to Thompson's (1967) conclusion that visual behaviors at least do not detract from effectiveness. (More is said about the congruency and focusing issues in the discussion of combination behaviors below.)

Movement. Also contrary to many of the findings of research in speech, communication, and education, no measure of teacher movement was significantly related to level of effectiveness in this study. Beyond some rather high positive and negative correlation coefficients, about all that can be safely stated is that teacher movement apparently does not interfere with effectiveness.

However, one notable consistency does exist in the June subsample, in that 12 of 13 measures of teacher movement are negatively correlated with effectiveness, five of the 13 correlates being significant at the .05 level and three of the 13 at the .10 level. In the August subsample the direction is reversed, although the level of significance is not as high. This could indicate that the less effective teachers in June were more "nervous," and that this nervousness manifested itself in a great deal of random movement.

Facial expression. Only one behavior achieved consistent results in this class, namely, congruent facial expressions. The interpretation of this finding must be tempered by the fact that congruency was determined by the frequency of occurrence of the selected verbal behaviors, the majority of which were negatively correlated with effectiveness. The negative correlation of congruent facial expressions is predictable on the basis of the high negative correlation with success of many of the verbal behaviors used in the determination of congruency.

Vocal behavior. One measure of vocal proficiency, vocal intensity, showed a clear, positive, and significant relationship with success in teaching for these 32 teachers. As in the case of movement, there was probably an element of "nervousness" in the June subsample of interns that should not be overlooked. In that sample, the less effective teachers had lower ratings on the use of voice than did the better teachers, which in turn accounted for much of the variance between the high- and low-scoring teachers in that sample. In August, however, this variable, although still positively correlated with success, was not as discriminating as it

was in June. The significance of this behavior in the combined group was great enough, however, to warrant the conclusion that vocal intensity, as defined in this study, was related to success in the teaching conditions encountered by the 32 teachers.

The results go in the same direction for total vocal rating as well, although the level of significance was only .10. Both results, however, attested to the value of vocal proficiency in effective teaching in this study.

Combinations of Behaviors

Many combinations of behaviors were investigated in an attempt to isolate meaningful patterns or relationships between the various behaviors. As implied in earlier parts of this report, this attempt yielded some interesting results, as summarized in Table 3.

Although seven measures of certain nonverbal behaviors in combination with selected verbal behaviors yielded results significant at the .05 level, the real significance of the findings is affected by factors already discussed under gestures. Furthermore, since the 31 different combinations measured were made up of behaviors which had, by themselves, some significant correlations with success, one could expect to find some combinations that achieve significant correlations as well.

The interesting results derive from the patterns that developed among certain combinations. For example, three verbal behaviors, i.e., explaining links, conditional words, and probing, when used with nonfocused gestures, produced negative correlations with effectiveness when measured in the combined sample. On the other hand, repetition and reinforcement, when each is used with a focused gesture, were positively correlated with effectiveness.

TABLE 3

Mean, Standard Deviation, Correlation With Twice-Adjusted Achievement Scores, Analysis of Variance, and Sign Test Statistics for all Significant Combination Behaviors

Behavior by Sample	Mean H	Mean L	SD H	SD L	rho	F	Sign[a]
1. Explaining Links Used with Nonfocused Gestures							
June Subsample (N=16)	.20	.30	.12	.15	-.36	2.0	
August Subsample (N=16)	.15	.20	.10	.11	-.26	1.0	
Total Sample (N=32)	.18	.25	.11	.14	-.31	2.8	NS
2. Conditional Words Used with Nonfocused Gestures							
June Subsample (N=16)	.14	.35	.09	.20	-.56	7.4*	
August Subsample (N=16)	.16	.22	.15	.14	-.18	0.5	
Total Sample (N=32)	.15	.28	.12	.18	-.35	5.9*	.01
3. Probing Used with Nonfocused Gestures							
June Subsample (N=16)	.15	.27	.15	.17	-.39	2.2	
August Subsample (N=16)	.25	.39	.19	.13	-.42	2.8	
Total Sample (N=32)	.20	.33	.17	.16	-.39	4.7*	.09
4. Repetition Used with Focused Gestures							
June Subsample (N=16)	.35	.33	.30	.15	.03	0.0	
August Subsample (N=16)	.55	.24	.26	.20	.57	7.5*	
Total Sample (N=32)	.45	.28	.30	.18	.31	3.9	.09

TABLE 3 (Continued)

Behavior by Sample	Mean H	Mean L	SD H	SD L	rho	F	Sign[a]
5. Reinforcement Used with Focused Gestures							
June Subsample (N=16)	.38	.33	.32	.24	.08	0.1	
August Subsample (N=16)	.54	.32	.15	.17	.60	7.3*	
Total Sample (N=32)	.46	.33	.26	.20	.33	2.5	
6. Reinforcement Used with Facial Expressions							
June Subsample (N=16)	.04	.13	.10	.12	-.41	2.5	
August Subsample (N=16)	.06	.15	.06	.09	-.54	5.5*	
Total Sample (N=32)	.05	.14	.08	.11	-.47	7.5**	.09
7. Repetitions Used with Facial Expressions							
June Subsample (N=16)	.49	.63	.42	.47	-.35	1.9	
August Subsample (N=16)	1.13	.81	.30	.31	-.38	2.9	
Total Sample (N=32)	.81	.72	.48	.37	-.35	5.1*	.02

H = High-scoring

L = Low-scoring

* $p < .05$

** $p < .01$

[a] Within lessons (N = 10)

These two results are explained by the fact that the verbal behaviors in the first case had small, but negative correlations with effectiveness, as did the nonverbal behavior, nonfocused gestures. When the joint occurrence of the two is coded, the result is one of producing an even larger negative correlation with effectiveness.

The same explanation applies to the second case mentioned above, except that the nonsignificant positive correlations were larger when coded on the basis of joint occurrence. In effect, this says there may be a cumulative effect operating in the use of certain behaviors. Undesirable behaviors can accumulate to the point where the net negative effect is significantly related to certain measures of success in teaching. The same applies to desirable behaviors as well. Stated differently the finding is that the greater the number of undesirable behaviors a teacher possesses, the greater the probability of these accumulating in such a fashion as to become associated with a low level of effectiveness.

Results of Stepwise Discriminant and Stepwise Regression Analysis

In addition to analysis of variance, correlation, and sign test statistics, two additional statistics, stepwise discriminant analysis and stepwise regression analysis, were employed in this investigation. Results of these two analyses have been mentioned previously in this report, and the purpose of the following discussion is to elaborate upon their operation and the interpretation of their results.

Stepwise discriminant analysis. Stepwise discriminant analysis was used to answer one question, namely: Given the fact that variable "x" provides an acceptable level of discrimination between the two groups (high-scoring and low-scoring), how many of the teachers in each group would actually be classified as "high" or "low" on the basis of this one variable? Although analysis of variance might yield an F-ratio significant at the .001 level for a given variable, this can often be attributed to the fact that one subject in the

high group used that behavior three to four times more often than any other subject in any other group. Analysis of variance, although it provides some test of the differences between the two groups for a given variable, tells nothing about how many of the high group subjects actually scored higher on that variable than did the subjects in the low group.

Thus, stepwise discriminant analysis was employed as an extension of the two other statistical procedures, rank correlation and analysis of variance, in an attempt to determine the significance of each variable's contribution to the proper classification of high-scoring and low-scoring teachers. Furthermore, this procedure allowed the determination of those two or three variables which best account for the variance between the two groups, i.e., best classify the cases in each group.

This analysis attempts, in stepwise fashion, to classify all high cases as "high" and all low cases as "low" in the following manner:

1. An F-ratio is computed for each variable, which is the same as an analysis of variance between high and low teachers for each variable.
2. The variable with the highest F-ratio, say A, is entered at Step 1, and all cases are ranked according to increasing magnitude on A, with those in the top half classified as high teachers (assuming a positive correlation of A with achievement.
3. For each remaining variable, V, a linear combination of A and V is computed which produces the largest possible F-ratio for high versus low.
4. The second variable entered, B, is the variable whose linear combination with A produces the highest F-ratio.

5. Then all teachers are ranked on the magnitude of the combined variables (the best linear function of A and B), and the top cases are "classified" as high.

6. The procedure is repeated for N steps.

Of course, variables A and B may not be the best combination of variables, but A is the best variable, and given that A is used, B is the best variable to add to it.

As can be seen in Tables 4-6, the additional variables at each step aid classification of each case in each of the two groups. However, as the following discussion points out, one must exercise caution when interpreting the results of this statistic.

First, only the first two variables entered can be interpreted as having any practical significance for the type of investigation undertaken here. This is due to the fact that after the first two steps most of the cases in each group have been properly classified, and variables entered in subsequent steps are selected on the basis of their ability to classify only the remaining misclassified cases. Thus, a variable which properly classified one high case and one low case will be entered, although the original analysis of variance for that variable may have yielded an F-ratio of no significance. As shown in Tables 4-6, most of the cases in each group have been properly classified on the basis of the first two variables, and the perfect classification of all cases occurs only when the analysis produces a variable that can classify the few remaining cases in each group. This occurs at Step 3 in the June sample, and Step 4 in the August sample, but fails to occur at any of the 16 steps in the combined sample.

Second, proper classification of all cases is more difficult when most variables considered have analysis of variance F-ratios significant at the .05 level, which is the case here. Those variables which achieve F-ratios significant at the .05 level tend to occur more frequently in the same high cases or less frequently in the same

Table 4

Summary of Stepwise Discriminant Analysis of the 22 Most Significant Variables in the June Subsample (N=16)

Step	Variable Entered	F-Ratio	F-Ratio at Entry		Cases Classified at Each Step	
					High	Low
1.	Verbal Markers of Importance	7.9	7.9	High	4	4
				Low	2	6
					High	Low
2.	Percent of Times Conditional Words Used with Nonfocused Gestures	7.4	5.1	High	7	1
				Low	2	6
					High	Low
3.	Percent of Times Verbal Reinforcement Used with Facial Expressions	2.5	8.5	High	8	0
				Low	0	8

TABLE 5

Summary of Stepwise Discriminant Analysis of the 22 Most Significant Variables in the August Subsample (N=16)

Step	Variable Entered	F-Ratio	F-Ratio at Entry		Cases Classified at Each Step: High	Low
1.	Percent of Nonreinforced Responses	9.6	9.6	High	6	2
				Low	1	7
2.	Total Frequency of Lateral and Forward-Backward Moves	3.1	4.4	High	6	2
				Low	1	7
3.	Vocal Intensity	1.7	4.4	High	7	1
				Low	1	7
4.	Total Congruent Facial Expressions	5.1	11.5	High	8	0
				Low	0	8

TABLE 6

Summary of Stepwise Discriminant Analysis of the 22 Most Significant Variables in the Total Sample (N=32)

Step	Variable Entered	F-Ratio	F-Ratio at Entry		Cases Classified at Each Step	
					High	Low
1.	Verbal Markers of Importance	10.6	10.6	High	9	7
				Low	3	13
					High	Low
2.	Vocal Intensity	7.6	6.5	High	13	3
				Low	2	14
					High	Low
3.	Inquiry-Observation Questions	0.6	4.8	High	14	2
				Low	2	14
					High	Low
4.	Total Nonfocused Gestures	3.8	5.2	High	14	2
				Low	2	14

TABLE 6 (Continued)

Step	Variable Entered	F-Ratio	F-Ratio at Entry		Cases Classified at Each Step	
					High	Low
5.	Percent of Times Verbal Marker of Importance Used with Distributed or Massed Repetitions	9.5	4.1	High	15	1
				Low	4	12
					High	Low
6.	Percent of Times Repetition Used with Focused Gestures	3.9	1.9	High	16	0
				Low	2	14
					High	Low
7.	Percent of Times Verbal Reinforcement Used with Facial Expressions	7.5	2.1	High	15	1
				Low	2	14

low cases, i.e., those teachers high in the use of one significant variable tend also to be high in the use of all other significant variables, as shown in Tables 4-6.

One additional indication of the usefulness of this form of analysis is gained by looking at its results for the most significant variable as originally determined through analysis of variance, verbal markers of importance. After computing stepwise discriminant analysis, as in Tables 4-6, it is quite obvious that the significance of this variable comes from the fact that it occurs infrequently in most of the low cases, not from the fact that it occurs frequently in the high group cases. Thus, of the 16 teachers in the low-scoring group, 13 actually scored low on that variable, whereas in the high group, only nine actually scored high.

This low frequency in the low-scoring group is further substantiated by observing the results for the June subsample, Table 4, where verbal markers of importance classified only one-half of the high cases, but classified three-fourths of the low cases. Also, in the August subsample, Table 5, verbal markers of importance is not entered at any of the first five steps, because most of the between-groups variance is accounted for by other variables, namely percent of nonreinforced pupil responses.

In summary, this form of analysis tends to improve the determination of practical significance by indicating the source as well as the amount of between-groups variance. One might well say that, given similar data for a different group of 32 teachers, it is likely that the behaviors included in the first three or four steps of stepwise discriminant analysis will properly classify the high-scoring and low-scoring teachers.

Stepwise Regression Analysis

This method was used in the analysis of the 22 most significant variables in the total sample. Much like stepwise discriminant analysis, this method attempts to locate that variable which contributes to the multiple correlation coefficient at each step in the analysis.

When used to analyze these 22 variables in the total sample, the four variables listed in Steps 1-4 in Table 7 were included before the F-ratio at entry for the fifth variable failed to achieve the .05 level of significance. The multiple correlation coefficient for the first four variables is .77, and these same variables are included in Steps 1-7 in stepwise discriminant analysis as shown in Table 6. Thus, the results of stepwise regression analysis and stepwise discriminant analysis show that the variables included in these analyses account for most of the differences between the high- and low-scoring teachers.

Summary

Close inspection of the results obtained through the various analyses conducted tends to indicate that success in the teaching tasks for the June subsample teachers is related to different variables than for the August subsample, and that these differences result in the identification of very few discriminating variables when the two samples are combined. Following is an attempt to summarize these data further by putting forth a description of the differences between the high-scoring and low-scoring teachers in each of the two samples analyzed.

The June Subsample

The differences in the June subsample can best be described by the fact that the high-scoring teachers tended to use fewer behaviors which reduce the chances of success. That is, the most consistent result is that the majority of the variables were negatively correlated with success in this subsample.

Although verbal markers of importance achieved the highest F-ratio, stepwise discriminant analysis indicates that not all high-scoring teachers used the behavior to a great degree, although 75 percent of the low-scoring teachers used it even less. Most of the high-scoring cases are so classified through their non-use of many other behaviors. Thus, it appears as if the high-scoring

TABLE 7

Summary of Stepwise Regression Analysis for the Combined June and August Samples (N=32)

Step	Variable Entered	rho at Step 0 (b)	rho at Entry	Multiple rho	F-ratio at Step 0	F-ratio at Entry
1.	Number of Distributed Repetitions Used with Verbal Markers of Importance	.48	.48	.48	8.9**	8.9**
2.	Vocal Intensity	.44	.46	.63	9.4**	7.9**
3.	Total Nonfocused and Congruent Gestures	-.34	-.48	.73	10.7**	8.4**
4.	Reinforcements Used with Facial Expressions	-.47	-.33	.77	9.5**	3.3*
5.	Repetitions Used with Focused Gestures	.31	.28	.79	8.5**	2.25

*p < .05
**p < .01

teachers were more successful for what they did not do than for what they did. When confronted with their first teaching situation, the high-scoring teachers tended to present rather "straightforward," "simple," and "to the point" kinds of lessons, avoiding "discussion" with students as well as the use of any significant number of nonverbal behaviors.

Also to be considered is the notion of "nervousness" discussed earlier. The consistent negative correlations of gestures and movements with success, and the positive correlations of voice with success seem fairly strong evidence that the lower-scoring teachers had less control over their nonverbal behavior, and that much of this uncontrolled behavior interfered with success on the teaching task. It would seem that the consistently positive relationship between these two variables and effectiveness in explaining reported by Rosenshine (1968), in which he analyzed the behavior of experienced social studies teachers, are not generalizable to preinterns, i.e., untrained and inexperienced teacher trainees. Once can hypothesize that each teacher investigated by Rosenshine was using a rather well-established repertoire of behaviors, derived from both training and experience. Preinterns, on the other hand, have no such repertoire to rely on, and it is possible that the low-scoring teachers exhibited a wide variety of nonverbal and verbal behaviors in random and uncontrolled ways.

The August Subsample

All of the lessons taught by teachers in the August subsample are similar in that all exhibited a great deal of teacher-student interaction. This is not surprising in light of the emphasis placed upon such interaction during the summer training program. Since variances in interaction are not related to effectiveness in any consistent way, one can surmise that the behaviors associated with high pupil-verbal involvement are irrelevant or they are not always used in an effective manner.

Also to be considered is the change in the mean frequency and the sign of the correlation coefficients for the various measures of teacher movement. First, 12 of the 13 measures of movement in June showed negative correlation coefficients, and all measures of movement in August showed positive correlation coefficients. Second, while both high-and low-scoring teachers in August exhibited more use of movement than their June counterparts, the increase for the low-scoring August teachers was only 30 percent.

Again, there are most likely some training effects reflected in these changes since the summer training program did include intensive instruction and practice in using various nonverbal behaviors designed to vary the stimulus situation within the classroom. One can hypothesize that the high-scoring August teachers were able to use more of these behaviors and use them more effectively than their low-scoring counterparts who experienced the same training. If one can accept the assumption that the August teachers had gained enough training and experience to approximate the effectiveness of the experienced teachers investigated by Rosenshine (1968), then his findings are reasonably well substantiated by the findings of this investigation.

The significance of verbal markers of importance in the June subsample, and percent of nonreinforced pupil responses and percent of multiply reinforced pupil responses in the August subsample, would seem to indicate that in both situations, the more effective teachers were able, through different behaviors utilized within different strategies (informal lecture versus high teacher-student interaction), to emphasize the major points of the lesson.

The Total Sample

Characterizing the behavioral differences between all high-scoring and all low-scoring teachers is more speculative than that of either the June or August subsamples.

The most defensible conclusion is the one presented earlier: that no matter what strategy is employed, in contrast to the low-scoring teachers, the high-scoring teachers in general did a better job of conveying the essential points of the lesson by emphasizing them through the use of repetition, verbal statements of importance, and reinforcement of pupil responses. These forms of verbal highlighting were supplemented by various forms of appropriately used nonverbal emphasis.

Little else can be inferred from these results. It must be remembered that these last few pages reflect conclusions reached by the investigator after a lengthy period of "dialogue" with the data. As such, they must be tempered by the restrictions inherent in this type of investigation.

Paucity of Significant Variables

It must be noted that none of the 100 or more measures of teacher behavior obtained in this investigation achieved F-ratios and correlation coefficients significant at the .05 level in the June, August, and combined samples. Although some consistently significant correlations were achieved, the search for single variables that discriminate all high-scoring from low-scoring teachers was unsuccessful.

One reason for this may have been the research design. Several controls were imposed upon the teachers in this study in order to increase the comparability of their performances on a common task. As discussed on pages 6-12, these controls included preset lessons and materials, teacher knowledge of the entire achievement test to be administered to the students, as well as standardized time allotment, room assignment, room furniture and room arrangement. In addition, the performance of each teacher was recorded on videotape by placing a mobile unit in each room in full view of the students and teacher. This addition may have suppressed some potentially disrupting student behaviors. In short, the research design may have assured all teachers in the study a degree of success not likely to have occurred had they been in a more "normal" teaching situation.

Finally, attempts were also made to remove any variance in mean class achievement scores that may have been due to differences in student ability or the difficulty of the 12 preset lessons used. It may be that the variables assumed to have been controlled are the only variables that actually account for most of the variance in teacher success, and that the variables investigated have only minimal effects.

A second possibility is that the search for variables which consistently discriminate all high-scoring from all low-scoring teachers is an unwise one, since few such variables may exist. As stated earlier, there may be an infinite number of discrete teacher or pupil behaviors that can fulfill the critical functions of effective classroom learning. If this is true, the design employed in this investigation is inappropriate since such variables are masked in the search for single variables which discriminate one heterogeneous group from another.

Two attempts were made to circumvent this problem. First, several combinations of verbal and nonverbal behavior were developed and investigated in order to determine the possible cumulative effect of certain behaviors. Despite the small proportion of such variables that were significant, the approach does seem promising. Secondly, stepwise discriminant analysis was employed in order to determine what groups of behaviors best account for the variance between the high- and low-scoring teachers, if there is no single variable that can do so. It would seem that further attempts to devise functionally related combinations of behaviors and the use of stepwise discriminant analysis to search for sources of variance may be fruitful. However, further knowledge of the precise effect of specific teacher behaviors upon the behavior of pupils is still lacking, and the need for cautious experimentation and replication is clear.

Implications for Teaching

Having been derived from a correlational study designed to discover promising correlates of effective teaching, these data cannot establish the causes of variance in teacher effectiveness. Although one might predict similar patterns of results for data collected under similar circumstances, extending such a prediction to "real" school situations is more than hazardous. As stated earlier, it would appear that the consistent use of one or more forms of emphasis or highlighting can contribute to the effectiveness of a single teacher presentation. Obviously, there is more to teaching than emphasis, but this investigator finds little in these results that provides any evidence of what this "more" might be.

Implications for Future Research

One could certainly build a case for replication of this investigation in order to establish the consistency of the findings. However, this investigator is of the opinion that continued correlational studies will not eliminate the possibility that the results obtained will be a function of one or more other unknown variables.

There is a need to assess the effect of systematically manipulated teacher behaviors on the subsequent behavior of pupils. That is, future studies should be experiments in which baseline data on pupil behavior is collected, environmental factors are brought under control, systematic treatments (teacher behaviors) are applied, and consequences are measured. It would seem that correlational studies contribute to this end by aiding in the identification of potential areas for experimentation. It is hoped that this investigation has made such a contribution, and that the wealth of data reported will lead others to fresh analyses and further investigation.

References

Ackerman, W. I. Teacher competence and pupil change. Harvard Educational Review, 1954, 24, 274-289.

Beighley, R. C. An experimental study of the effect of four speaking variables on listener comprehension. Speech Monographs, 1954, 19, 249-258.

Belgard, M., Rosenshine, V. B., & Gage, N. L. The teacher's effectiveness in explaining: Evidence on its generality and correlation with students' ratings. Research and Development Memorandum No. 10. Stanford Center for Research and Development in Teaching, 1967 (mimeographed).

Bellack, A. A., & Davitz, J. R. The language of the classroom. New York: Institute of Psychological Research, Teachers College, Columbia University, 1963.

Bellack, A. A., Kleibard, H. M., Hyman, R. T., & Smith, F. L. The language of the classroom. New York: Columbia University, 1966.

Berlyne, D. E. Conditions of pre-questioning and retention of meaningful verbal learning. Journal of Educational Psychology, 1966, 57, 128-132.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Coats, W. D., & Smidchens, U. Audience recall as a function of speaker dynamism. Journal of Educational Psychology, 1966, 57, 189-191.

Flanders, N. A. Teacher influence, pupil attitudes, and achievement. Minneapolis: College of Education, University of Minnesota, 1960. (Final Report, Cooperative Research Project No. 397, U. S. Office of Education.)

Fortune, J., Gage, N. L., & Shutes, R. E. A study of the ability to explain. Paper presented at the meeting of the American Educational Research Association, Chicago, February 1966.

Gage, N. L. Paradigms for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Gallagher, J. J., & Aschner, M. J. A preliminary report: Analysis of classroom interaction. Merrill-Palmer Quarterly of Behavior and Development, 1963, 9, 183-194.

Medley, D. M., & Mitzel, H. E. Some behavioral correlates of teacher effectiveness. Journal of Educational Psychology, 1959, 50, 239-246.

Medley, D. M., & Mitzel, H. E. Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Mitzel, H. E. A behavioral approach to the assessment of teacher effectiveness. New York: Division of Teacher Education, College of the City of New York, 1957 (mimeographed).

Mitzel, H. E., & Gross, C. Development of pupil growth criteria in studies of teacher effectiveness. Educational Research Bulletin, 1958, 37, 178-187.

Petrie, C. R., Jr. Informative speaking: A summary and bibliography of related research. Speech Monographs, 1963, 30, 79-92.

Rosenshine, V. B. Behavioral correlates of teacher effectiveness in explaining social studies material. Unpublished doctoral dissertation, Stanford University, 1968.

Rudin, S. A. Measuring the teacher's effectiveness as a lecturer. Journal of Genetic Psychology, 1961, 98, 147-154.

Ryans, D. G. Characteristics of teachers: Their description, comparison, and appraisal. Washington, D. C.: American Council on Education, 1960.

Smith, B. O. A conceptual analysis of instructional behavior. Journal of Teacher Education, 1963, 14, 294-298.

Smith, B. O., & Meux, M. O. A study of the logic of teaching. U. S. Department of Health, Education, and Welfare. Project No. 258, Urbana, Illinois: University of Illinois, 1962.

Swift, L. F. Explanation. In B. O. Smith and R. H. Ennis (Eds.), Language and concepts in education. Chicago: Rand McNally, 1961.

Taba, H. Thinking in elementary school children. Cooperative Research Project No. 1574. Washington, D. C.: Office of Education, 1964.

A4

Thompson, W. N. Quantitative research in public address and communication. New York: Random House, 1967.

Thyne, J. M. The psychology of learning and techniques of teaching. London: University of London Press, 1963.